

Aplicación del Análisis Discriminante y Regresión Logística en el estudio de la morosidad en las entidades financieras. Comparación de resultados

M. Jesús Mures Quintana
Ana García Gallego
M. Eva Vallejo Pascual

En este trabajo se presentan los resultados de un estudio empírico realizado en Castilla y León, mediante la aplicación de dos técnicas estadísticas a una muestra de clientes de entidades financieras implantadas en dicha Comunidad, con el fin de valorar el riesgo de crédito. A su vez, se determina el método que permite discriminar mejor entre clientes morosos y no morosos, lo que se realiza a partir de una serie de factores que influyen en el comportamiento de pago de los clientes.

Palabras clave: riesgo de crédito, morosidad, análisis discriminante, regresión logística.

In this paper we present the results of an empirical study of the region of Castilla & León (Spain), through the application of two statistical techniques to a sample of costumers of financial institutions settled in the said region, in order to evaluate the credit risk. On the other hand, we determine the method that better discriminates between default and non-default costumers, what is obtained through a series of factors that influence on their payment behaviour.

Key words: credit risk, default, discriminant analysis, logistic regression.

1. INTRODUCCIÓN

La incertidumbre que caracteriza el entorno económico actual al que deben enfrentarse las empresas en el desarrollo de su actividad ha

convertido el riesgo en un elemento esencial que han de descubrir, analizar y valorar de forma adecuada. Entendiéndose por riesgo, en términos generales, la variabilidad del beneficio esperado o pérdida de valor económico, en el ámbito de la empresa se pueden encontrar distintos tipos de riesgo, que se concretan en dos principales: el riesgo económico, que afecta a los resultados de explotación de la empresa, y el riesgo financiero, resultado del diseño de la estructura de capital de la misma.

En el caso de las entidades financieras o de crédito, a los riesgos anteriores, que no les son ajenos, se añaden un conjunto de riesgos adicionales que podríamos calificar como propios de su actividad, y que se sintetizan en cinco grandes apartados: riesgo de mercado, riesgo de crédito o insolvencia, riesgo de liquidez, riesgo operacional y de tecnología, y riesgo legal.

De todos estos riesgos, el riesgo de crédito es el típico de la actividad financiera y el que puede provocar mayores pérdidas potenciales, de ahí que sea el que mayor número de crisis financieras ha provocado y al que se dedica mayor atención.

Sin embargo, no siempre ha sido así, pues los gestores empresariales se habían caracterizado por tomar la mayoría de decisiones con el objetivo de evitar o, por lo menos, minimizar cualquier tipo de riesgo, dada la connotación claramente negativa que aquél tenía en todas las actividades económicas. Esto era posible por la existencia de una reducida competitividad, economías cerradas y fuerte proteccionismo de los estados a sus empresas, si bien, en la medida en que las economías se han ido haciendo cada vez más abiertas y competitivas, esta estrategia de aversión al riesgo, característica de la gestión empresarial tradicional, se ha hecho imposible. De esta forma, el riesgo ha pasado a ser un factor esencial que hay que gestionar adecuadamente para obtener una rentabilidad acorde con el nivel de riesgo asumido, constituyendo ambos elementos, junto con la solvencia, los tres ejes de la gestión en el ámbito de las empresas, y, especialmente, en las entidades financieras.

Si nos centramos en el riesgo de crédito, las técnicas y modelos utilizados en su gestión han seguido una evolución acorde con la exigencia, cada vez mayor, de los gestores empresariales demandantes de herramientas para la toma de decisiones en la asunción de riesgos. Aunque en sus inicios los modelos de valoración del riesgo de crédito eran de corte univariante, contemplando el riesgo crediticio en función de una sola variable, o de considerar varias, analizando cada una por separado, el

hecho de que la asunción de riesgos dependa de múltiples factores, provocó que las técnicas evolucionaran hacia modelos de tipo multivariante. Dentro de éstos, distinguimos, en primer lugar, aquéllos de carácter subjetivo, basados en el criterio y juicio de analistas expertos, quienes utilizan sus conocimientos y experiencias acumuladas a través de años de trabajo para evaluar el riesgo de crédito y tomar decisiones al respecto. Con el fin de superar los inconvenientes que tales modelos presentan, en cuanto a la subjetividad de las decisiones, se desarrollaron después modelos objetivos de evaluación, basados en métodos estadísticos, que consiguen que distintos analistas, a partir de idénticos elementos o factores, lleguen a las mismas decisiones.

Todas las herramientas cuantitativas de medición del riesgo de crédito descansan sobre dos piezas básicas, el *scoring* y el *rating*, que permiten calificar el riesgo mediante una puntuación o una clasificación en grupos de riesgo, y que utilizan técnicas estadísticas como el análisis discriminante o los modelos de probabilidad condicionada, entre los que se incluyen el *probit* y el *logit*.

En este contexto, se plantea este trabajo, cuyo objetivo fundamental se centra en el análisis y evaluación de la morosidad, como forma de manifestación del riesgo de crédito, mediante la determinación de aquellos factores con mayor influencia en dicho fenómeno. Esto se realiza a través de un estudio comparativo de dos técnicas estadísticas adecuadas para este análisis, como son el análisis discriminante, tradicionalmente utilizado por las entidades financieras en sus *credit scoring*, y la regresión logística, que presenta ciertas ventajas respecto al anterior, aplicándose ambas a una muestra de clientes de tres tipos de entidades financieras con implantación en la Comunidad de Castilla y León.

Para ello, el presente trabajo comienza con un apartado dedicado al riesgo en las entidades financieras, con especial referencia al de crédito, continuando con el desarrollo de la metodología aplicada en el estudio empírico, que recoge una breve exposición de los aspectos teóricos en que se sustentan las dos técnicas estadísticas aplicadas. A continuación se desarrolla el análisis empírico, que comprende una serie de etapas que van desde la determinación de aquellas variables que permiten explicar la morosidad, hasta el procedimiento de selección de la muestra de clientes objeto de estudio, para finalizar con la exposición de los principales resultados obtenidos.

2. EL RIESGO EN LAS ENTIDADES FINANCIERAS

2.1. *Concepto y tipos de riesgo*

La actividad económica se desarrolla en un ambiente de incertidumbre, convirtiendo el riesgo en un factor inherente a la misma. De este modo, surge aquél como la contingencia, probabilidad o proximidad de un daño o peligro, en concreto, de sufrir una pérdida. La incertidumbre, junto con la aleatoriedad, constituyen las características principales del riesgo, añadiéndose como tal el conflicto, ya que el riesgo se presenta ante situaciones diferenciadas entre las que elegir.

Así pues, el riesgo se ha convertido en uno de los rasgos básicos del entorno económico actual al que se enfrentan las empresas, que deben asumir riesgos en su toma de decisiones, en busca de la máxima rentabilidad, hablándose del binomio riesgo-rentabilidad como algo inseparable de la gestión de las mismas. En el caso de las entidades financieras, esta característica es esencial a la actividad que desarrollan, consistente en la concesión de créditos, asumiendo un riesgo cuando prestan unos recursos financieros que otros clientes les han cedido, sin controlar posteriormente el destino y utilización de los mismos.

El riesgo puede definirse, de forma más completa, como la exposición a la posibilidad de pérdidas económicas y financieras, de daños en las cosas y en las personas o de otros perjuicios como consecuencia de la incertidumbre que origina el llevar a cabo una determinada acción. Esta incertidumbre que, como venimos comentando, es una de las características fundamentales del riesgo, se presenta en tres aspectos:

- en cuanto a la posibilidad de que se materialice una pérdida efectiva
- sobre la cuantía de la pérdida
- sobre el instante en que puede materializarse dicha pérdida.

El riesgo global en la actividad financiera resulta de la suma de distintos tipos de riesgos, que se describen, de forma breve, a continuación:

a) Riesgo de mercado: supone la probabilidad de pérdida ante movimientos adversos en los precios de los instrumentos financieros en los mercados donde se tengan posiciones. En función de su causa, se distingue, a su vez: riesgo de posiciones en divisas, riesgo de tipo de

interés/precio, riesgo de operaciones con opciones y otros riesgos de mercado, relativos a correlación, pago anticipado y aseguramiento.

b) Riesgo de crédito: corresponde al riesgo de pérdida que se produce ante el incumplimiento de los pagos adeudados a la entidad, presentando como modalidades: riesgo de contraparte o sustitución, riesgo país, riesgo de liquidación o entrega y riesgo de emisor.

c) Riesgo operacional y de tecnología: comprende un conjunto de riesgos que surgen por la posible ocurrencia de sucesos inesperados, relacionados con la infraestructura operativa y tecnológica, tanto interna como externa, de la entidad.

d) Riesgo de liquidez: entendiéndose por liquidez la capacidad de transformar un activo en efectivo a los precios existentes en cada momento, sin incurrir en más pérdidas que las que, en su caso, imponga el mercado, puede definirse este riesgo como la posibilidad de que una entidad financiera no disponga de liquidez y no pueda hacer frente a sus compromisos vencidos, ya sea por la imposibilidad de modificar una posición a tiempo en el mercado en un momento dado, por un desfase temporal en los flujos de caja o por necesidades de liquidez no previstas.

e) Riesgo legal: incluye una variedad diversa de riesgos que se generan por la imposibilidad de ejecutar un contrato en los términos previstos, y que comprenden el riesgo de documentación, el de legislación específica de cada país y el de capacidad de las contrapartes.

2.2. Riesgo de crédito

De los distintos tipos de riesgo que acabamos de describir, puede considerarse el riesgo de crédito como el más importante al que deben hacer frente las entidades financieras, por ser intrínseco a la actividad que desarrollan, y porque es la principal incertidumbre a la que estas entidades se enfrentan en las operaciones de activo que les vinculan a sus clientes, de ahí que este trabajo se centre en ese riesgo.

El riesgo de crédito podría definirse, de forma genérica, como el posible quebranto que originaría a la entidad una variación en las condiciones y características de una contrapartida, que pudiese alterar la capacidad de ésta para cumplir con los términos contractuales de una operación.

Una definición alternativa, que recoge las dos formas de manifestación del riesgo de crédito, es la que lo describe como la incertidumbre derivada de la posibilidad de sufrir un quebranto por el incumplimiento de alguna o de todas las obligaciones contractuales de la contraparte o contrapartida en una operación financiera, ya sea por entrada en *mora* del deudor, provocada por el retraso en dicho cumplimiento, o por el impago definitivo de sus obligaciones, lo que deviene en la *insolvencia* del mismo.

Esta distinción entre las manifestaciones del riesgo crediticio aparece, asimismo, recogida en diversas Circulares del Banco de España, que clasifican el crédito respecto a su situación, es decir, el estado de cumplimiento de las obligaciones de pago correspondientes. Atendiendo a esta normativa, se consideran como morosos aquellos créditos y demás operaciones vencidos y no cobrados, sin mediar novación o prórroga, cuando hayan transcurrido más de tres meses desde su vencimiento. Sin embargo, con carácter general, se emplea el calificativo de moroso para designar a la persona o deudor que se retrasa en el pago de sus deudas o en el cumplimiento de su obligación, al no efectuarlo en la fecha de vencimiento estipulado.

El riesgo de crédito se clasifica en dos grandes tipos de riesgos: el riesgo de contrapartida, cuyo origen está en el cambio de solvencia de una contrapartida o conjunto diferenciado de ellas, y el riesgo país, que se corresponde con la solvencia de contrapartidas que pertenecen a un área geográfica política y legalmente definida como estado. No obstante, dentro de cada uno de estos riesgos pueden establecerse, a su vez, otras clasificaciones, según distintos criterios, como el evento o factor que genera el riesgo, la operación en que se manifiesta, el grado de concentración del riesgo o su estimación.

Por otro lado, el factor riesgo que acompaña a toda operación que las entidades formalizan con sus clientes, con el fin de concederles crédito, aparece en tres momentos clave de la relación con el cliente:

- En el estudio de la operación: la entidad financiera tiene que realizar un análisis riguroso del cliente antes de tomar la decisión, por lo que dedica un cierto tiempo a la evaluación de la operación, a partir de la documentación e información necesaria.
- En la formalización de la operación: se debe actuar de forma rigurosa, estableciendo los controles de seguridad adecuados.

- Entre la formalización y la cancelación: durante la vida de la operación, la entidad financiera debe realizar el oportuno seguimiento de la misma y del cliente.

Además, para medir el nivel de riesgo asociado a una operación de activo, es necesario considerar una serie de factores determinantes del mismo, que se refieren, en primer lugar, a las características de la operación. Dentro de éstas, se incluyen el plazo o vencimiento del término de la operación, el importe o cuantía, la modalidad o tipo de instrumento de financiación solicitado, y el destino o uso de la inversión financiada.

Otro factor que influye en el nivel de riesgo, relacionado con la operación, son las garantías aportadas por el cliente como medio requerido por las entidades financieras para disminuir el riesgo que conllevan las operaciones de activo.

Por último, el nivel de riesgo se mide a través del conocimiento del solicitante de la operación. Para llegar a ese conocimiento, las entidades financieras realizan un análisis riguroso del cliente, a través del estudio de información interna de origen histórico, obtenida de bases de datos de la entidad. A su vez, estudian información externa, solicitada al propio cliente o procedente de consultas a registros como el de Aceptaciones Impagadas (RAI), CIRBE, ASNEF, Registro de la Propiedad y Registro Mercantil.

3. METODOLOGÍA

Una vez presentados los riesgos a que deben enfrentarse las entidades financieras y, en especial, el de crédito, el presente trabajo se completa con un estudio empírico, cuyo objetivo es analizar y valorar la morosidad, como forma de manifestación de dicho riesgo, en las entidades financieras. Para ello, es preciso aplicar una metodología determinada.

De lo que se trata es de determinar los factores de mayor influencia en el comportamiento de pago de los clientes de las entidades financieras y que permiten distinguir los clientes solventes, que cumplen con sus obligaciones, de los morosos, que las incumplen o se retrasan en su cumplimiento. Por tanto, el objetivo del trabajo se centra en explicar el comportamiento de una variable categórica con dos modalidades: ser un cliente moroso o bien un cliente no moroso.

Dentro de los métodos estadísticos que pueden aplicarse en el análisis de este tipo de variables, tradicionalmente las entidades financieras han aplicado el análisis discriminante en los *credit scoring* que utilizan para el estudio de las operaciones de crédito que conceden a sus clientes, aunque los inconvenientes que tal técnica puede presentar han propiciado la aplicación de otras técnicas, como la regresión logística o *logit*.

Partiendo de las consideraciones anteriores, en este trabajo llevamos a cabo un estudio comparativo de los métodos de análisis discriminante y de regresión logística, a fin de determinar cuál de dichas técnicas, aplicables en el estudio y predicción del comportamiento de una variable de tipo cualitativo, permite realizar un mejor análisis del fenómeno de la morosidad en las entidades financieras. Por ello, dedicamos este epígrafe a exponer los aspectos teóricos más representativos de cada uno de los métodos aplicados. Ambos se engloban dentro de las técnicas de dependencia, cuyo objetivo es explicar o predecir un fenómeno, definido por una variable (o conjunto de variables) que actúa como dependiente, en función de una serie de factores que se relacionan con él, y que actúan como variables independientes o predictoras.

3.1. Análisis discriminante

- El análisis discriminante permite estudiar las diferencias entre dos (en el caso del análisis simple) o más (estaríamos ante el análisis discriminante múltiple) grupos de individuos definidos a priori, con respecto a varias variables simultáneamente.
- Tiene como objetivo analizar la relación entre una variable dependiente categórica con g modalidades, que se corresponden con los grupos analizados, y un conjunto de variables independientes $x_1, x_2 \dots x_p$, métricas o cuantitativas, a partir de una serie de funciones discriminantes, que son combinaciones lineales de las variables independientes que mejor discriminan o separan los grupos, y cuya expresión es la siguiente:

$$f_{km} = \mu_0 + \mu_1 x_{1km} + \mu_2 x_{2km} + \dots + \mu_p x_{pkm}$$

siendo f_{km} el valor o puntuación en la función discriminante para el caso m en el grupo k ; x_{ikm} el valor de la variable discriminante x_i para el caso m en el grupo k y μ_i los coeficientes o ponderaciones de las variables x_i .

- El número de funciones discriminantes a obtener depende, a su vez, del número de grupos definidos por la variable dependiente, ya que se obtienen tantas funciones como grupos menos uno, salvo que el número de variables independientes incluidas en el modelo sea inferior, en cuyo caso el número de funciones discriminantes obtenidas coincide con el de variables.
- Las funciones discriminantes se obtienen de forma que la primera contiene aquellas variables explicativas cuyos valores más diferencian los distintos grupos, la segunda función es la segunda combinación de variables que más discrimina entre los grupos, pero con la condición de que los valores obtenidos mediante la primera función no se hallen correlacionados con los de aquélla, y así sucesivamente.
- Estimadas las funciones discriminantes, su capacidad predictiva se evalúa mediante el establecimiento de una puntuación de corte óptima, que permite asignar los casos a cada uno de los grupos definidos por la variable dependiente, obteniéndose de la puntuación discriminante que corresponde a cada caso, a partir de los valores que presenta el individuo en la combinación de variables explicativas que forman las funciones discriminantes.

Dado que en nuestro trabajo se trata de diferenciar entre dos grupos de clientes (los solventes y los insolventes o morosos), se aplica el análisis discriminante simple, en el que la variable dependiente es dicotómica, siendo sus modalidades cada uno de los dos grupos definidos a priori, y, por tanto, se obtiene una única función discriminante. En cuanto a las variables independientes o predictoras que permiten explicar el comportamiento de la variable respuesta, este método impone la condición de que sean cuantitativas o métricas, aunque pueden introducirse también variables cualitativas o categóricas, sin más que transformándolas previamente en tantas variables ficticias o *dummy* como modalidades tenga la variable original menos una. Esta categoría actúa como referencia y aparece codificada con el valor cero en todas las variables ficticias, mientras que el resto de categorías se codifican con el valor uno en la variable *dummy* respectiva, si bien existen otras codificaciones. De esta forma, se facilita el tratamiento de las variables cualitativas como cuantitativas, interpretándose como el efecto en la variable dependiente cuando un individuo presenta determinada modalidad respecto a aquél que presenta la categoría de referencia.

3.2. Regresión logística

Esta técnica es similar a la anterior, aunque con ciertas diferencias que constituyen una serie de ventajas con respecto al análisis discriminante.

- La variable dependiente o respuesta también presenta dos categorías, pero en este caso representan la ocurrencia y no ocurrencia del acontecimiento definido por la variable, codificándose con los valores uno y cero, respectivamente. Por lo que se refiere a las variables independientes o explicativas, no se establece ninguna restricción, pudiendo ser cuantitativas, tanto continuas como discretas, y categóricas, con dos o más modalidades.
- Definida la variable dependiente como la ocurrencia o no de un acontecimiento, el modelo de regresión logística la expresa en términos de probabilidad, utilizando la función logística para estimar la probabilidad de que ocurra el acontecimiento o de que un individuo elija la opción uno de la variable dependiente, dados determinados valores de las variables explicativas, mediante la siguiente formulación:

$$\pi_i = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}}} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi})}} \quad , , \quad i = 1, 2 \dots n$$

siendo $\pi_i = P(y = 1)$.

Puesto que el modelo anterior no es lineal respecto a las variables independientes, se considera la inversa de la función logística, que es el *logit* o logaritmo de la *odds* o ventaja de que un suceso ocurra, definiéndose ésta como el cociente entre la probabilidad de que ocurra un acontecimiento y la probabilidad de que no ocurra, que es su complementaria, como puede observarse en la siguiente expresión:

$$\text{logit}(\pi_i) = \ln \left[\frac{\pi_i}{1 - \pi_i} \right] = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi} \quad , , \quad i = 1, 2 \dots n$$

La formulación anterior facilita la interpretación del modelo y de sus coeficientes, que reflejan, de este modo, el cambio en el *logit* correspondiente a un cambio unitario en la variable independiente considerada.

- La capacidad predictiva del modelo de regresión logística se valora mediante la comparación entre el grupo de pertenencia observado y

el pronosticado por el modelo, que clasifica a los individuos en cada grupo definido por la variable dependiente en función de un punto de corte establecido para las probabilidades predichas a partir de los coeficientes estimados y del valor que toman las variables explicativas para cada individuo.

4. ANÁLISIS EMPÍRICO

Como ya hemos dicho, el presente trabajo tiene como objetivo determinar los factores que caracterizan la morosidad o insolvencia en las entidades financieras, mediante la aplicación de las técnicas descritas en el apartado anterior a una muestra de clientes de entidades operativas en Castilla y León. Para ello, es preciso, en primer lugar, definir la variable dependiente que modeliza el fenómeno de la morosidad, así como determinar un conjunto de factores que influyen en el comportamiento de pago de los clientes, y que se consideran como variables independientes o explicativas de la morosidad. Un segundo aspecto a tener en cuenta es la selección de la muestra de clientes objeto del estudio, una vez fijadas previamente las entidades financieras de las que extraer dicha muestra. Este análisis se finaliza con la presentación de los principales resultados obtenidos.

4.1. Selección de variables

El primer aspecto a considerar para la realización del estudio empírico se refiere a la selección de las variables a incluir en el modelo a desarrollar, tanto la dependiente, que define el fenómeno cuyo comportamiento se trata de explicar, como las independientes o explicativas de dicho fenómeno.

En razón de los aspectos teóricos de las técnicas aplicadas en el estudio, la variable dependiente o respuesta que modeliza el fenómeno que se analiza es dicotómica, siendo sus modalidades la morosidad y no morosidad de la operación, que se han codificado, respectivamente, con los valores uno y cero.

Por lo que se refiere a las variables independientes o predictoras que permiten explicar el comportamiento de la variable dependiente, se han seleccionado teniendo en cuenta los factores que las entidades financieras consideran para medir el nivel de riesgo en las

operaciones de crédito que conceden a sus clientes. Como hemos señalado anteriormente, los mencionados factores hacen referencia a las características de la operación, incluidas las garantías aportadas, así como al conocimiento que se tenga del solicitante.

Respecto a la información que las entidades recogen a la hora de estudiar la concesión de un crédito, destacan tres elementos que resultan en dicha concesión: por un lado, la personalidad del solicitante, referente al carácter, integridad y naturaleza del individuo, que determina si pagará, o, por el contrario, existirá riesgo de crédito; por otro, la capacidad y el capital, relativos a la disponibilidad de ingresos para pagar y a la situación financiera y patrimonial del cliente, respectivamente, que establecen el límite de crédito que puede concederse, esto es, si el cliente podrá pagar y cuánto. Pero además de la información obtenida por las entidades sobre los factores personales, profesionales y patrimoniales de sus clientes, el hecho de que un solicitante de crédito puede hacer frente a sus obligaciones de pago también depende de las características de la operación que se formalice.

Por todo lo anterior, las variables independientes seleccionadas para su posible inclusión en el modelo se han clasificado en dos bloques:

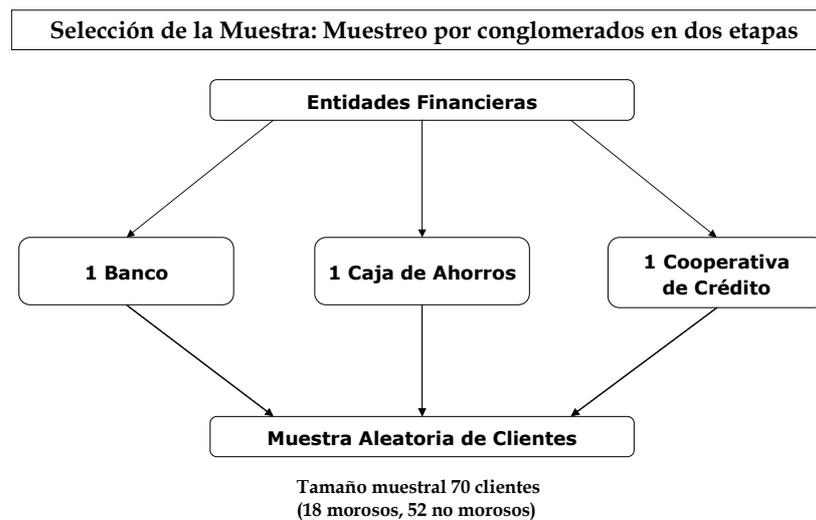
- Variables relativas a la operación de crédito o préstamo, tales como el importe solicitado, la duración o plazo de devolución del crédito, el destino de la inversión, la procedencia de la operación y las garantías aportadas.
- Variables relativas al perfil del solicitante, que incluyen aquéllas referidas a sus características personales, como edad, estado civil o cargas familiares; las que reflejan su situación profesional, como categoría laboral, ingresos mensuales o antigüedad en el trabajo, junto a otras que definen su relación con la entidad, como compensaciones ofrecidas, número de operaciones activas e importe total de las mismas o si tiene otras operaciones en las que se ha producido impago.

Cada una de estas variables, con sus respectivas categorías, se recoge en el anexo A1.

4.2. Muestra de entidades

Determinadas las posibles variables explicativas del comportamiento de pago de los clientes de entidades financieras, el siguiente paso en el análisis empírico es la selección de la muestra de clientes objeto del estudio.

Para formar la muestra, nos centramos en clientes particulares o personas físicas -excluyendo, por tanto, las personas jurídicas o sociedades-, considerando las operaciones de crédito o préstamo que han formalizado para la financiación de sus necesidades de consumo o de su actividad profesional. A fin de analizar el fenómeno de la morosidad, determinando las variables que permiten distinguir entre clientes morosos y no morosos, se seleccionan tanto clientes que han incumplido sus obligaciones de pago y se encuentran en situación de morosidad, como clientes que pagan sus deudas a su vencimiento y son solventes.



De la red de oficinas de las entidades financieras, entendiendo cada una como un conglomerado de clientes, se tomó una muestra total de nueve conglomerados (tres sucursales correspondientes a cada tipo de entidad), y dentro de cada oficina se extrajo una muestra aleatoria de clientes, entre todos aquéllos que tenían formalizada una operación de crédito o préstamo, fueran morosos o no.

El procedimiento de selección por azar dio lugar a una muestra total de 70 clientes, clasificados en 18 clientes morosos y 52

clientes que cumplen con sus obligaciones de pago, lo que supone un 26% de individuos morosos frente a un 74% de individuos no morosos, porcentajes éstos que difieren de los que se dan en algunas entidades financieras, que presentan valores más bajos para la proporción de clientes morosos.

Metodológicamente, existe justificación estadística a la selección de la muestra de clientes por la técnica de emparejamiento, evitando así que el problema de la diferencia del tamaño de ambos grupos de clientes, que es mucho más grande para los individuos no morosos frente a los morosos, se trasladara a la muestra si ésta se seleccionara por azar. Sin embargo, los inconvenientes que tal técnica presenta¹, y, sobre todo, la dificultad que el propio procedimiento conlleva de encontrar un cliente no moroso por cada moroso seleccionado con los mismos valores para cada variable independiente incluida en el modelo, han llevado a la extracción de la muestra de forma aleatoria.

4.3. Análisis de Resultados

Una vez seleccionada la muestra de clientes, se procede a su análisis, mediante la aplicación de las técnicas de análisis discriminante y de regresión logística, con el objetivo de determinar las variables más significativas en la predicción de la morosidad.

El estudio de los datos correspondientes a los clientes que forman la muestra puede realizarse desde una perspectiva univariante, aplicando ambos métodos a cada variable independiente seleccionada para su posible inclusión en el modelo. Aunque este análisis unidimensional permite determinar la relación entre cada variable explicativa y la explicada, el sentido de la misma y su significación, en todo fenómeno económico, incluido el comportamiento de pago, y por consiguiente, la morosidad de un sujeto, influye más de una variable o factor. Por ello, es preciso estimar modelos multivariantes, en los que la variable respuesta viene explicada por un conjunto de variables predictoras.

La selección inicial de variables explicativas del fenómeno analizado, realizada sobre la base de la teoría que sustenta cada campo científico, puede ser muy grande, por lo que en la mayoría de estudios se

¹ Palepu (1986) realiza una comparación entre la selección aleatoria de la muestra y la selección por emparejamiento, señalando los inconvenientes más importantes que ésta presenta.

busca un modelo parsimonioso, que permita explicar la mayor cantidad posible de variabilidad de la variable dependiente con un conjunto mínimo de variables que contribuyan de forma significativa a esa explicación.

Con el fin de encontrar el número óptimo de variables independientes a incluir en el modelo pueden aplicarse distintos métodos de selección de variables, que son de tipo secuencial y que se clasifican en tres tipos: métodos de selección hacia delante o *forward*, métodos hacia atrás o *backward* y métodos por pasos sucesivos o *stepwise*. Los primeros comienzan con el modelo que incluye sólo el término constante y se van añadiendo al mismo variables independientes según su grado de relación con la variable dependiente y su significación estadística. El procedimiento contrario siguen los segundos, pues el análisis comienza con todas las variables explicativas incluidas en el modelo, y se procede a la eliminación de las variables cuya aportación a la explicación de la variable dependiente no sea significativa. Por último, la selección de variables por pasos sucesivos es una combinación de los dos métodos anteriores; como en los métodos hacia delante, partiendo del modelo inicial, se van introduciendo variables en el mismo, pero además, en cada paso, existe la posibilidad de eliminar una variable ya incluida en la ecuación, si deja de ser significativa en la explicación de la variable respuesta.

De acuerdo con todo lo expuesto, siendo el objetivo del estudio encontrar un modelo con el menor número posible de variables y la máxima capacidad predictiva, se ha considerado, tanto en la aplicación del análisis discriminante como de la regresión logística, un procedimiento de selección por pasos sucesivos o *stepwise selection*, cuyos resultados se muestran a continuación.

4.3.1. Resultados del análisis discriminante

Las funciones discriminantes, obtenidas como combinaciones lineales de las variables explicativas en el análisis discriminante, permiten clasificar a los individuos de la muestra en los grupos definidos por la variable dependiente, a través del establecimiento de un punto de corte para las puntuaciones calculadas a partir de la función correspondiente.

En este caso, se trata de una variable respuesta con dos modalidades (los dos grupos de clientes, morosos y no morosos), por lo que se obtiene una única función discriminante. En el programa utilizado en el estudio empírico (SPSS), son varios los criterios que pueden considerarse

en la selección de las variables que se incluyen en las funciones discriminantes: Lambda de Wilks, varianza no explicada o residual, distancia de Mahalanobis, V de Rao y menor razón F. En nuestro estudio hemos aplicado el criterio de minimización de la lambda de Wilks para la presentación de los resultados, aunque coinciden para el resto de criterios que pueden utilizarse.

- La función discriminante estimada para los clientes de la muestra contiene cinco variables independientes, cuyas ponderaciones, sin tipificar y estandarizadas, así como los estadísticos que valoran su significación individual, se recogen en la tabla 1:

Tabla 1. Función discriminante estimada

Variable	Coefficientes f. discriminante	Coefficientes estandarizados	Lambda de Wilks	F exacta	Sig.
NUMIMPAG	1,704	1,013	,502	67,352	,000
NUEVARES	3,856	,624	,348	41,249	,000
DURARETR	,125	,567	,419	46,528	,000
GARHIPOT	-1,031	-,502	,301	37,687	,000
ESTCIVIL	-,611	-,440	,266	35,290	,000
Constante	,779				

Dos de las cinco variables incluidas en la función discriminante hacen referencia al perfil del solicitante, en concreto, a su estado civil y si es de nueva residencia en la zona. También ha mostrado ser significativa la aportación de garantía hipotecaria o real, que resulta ser una característica de la operación formalizada. Las dos últimas variables, relativas al número de impagos anteriores y a la duración del retraso correspondiente, si bien se refieren a la operación, también guardan relación con el cliente, por cuanto que son indicadores de su personalidad y naturaleza respecto al pago de sus deudas; en realidad, podrían considerarse como variables que definen la relación entre cliente y entidad en cuanto a las operaciones de crédito o préstamo que han formalizado.

Por lo que se refiere al signo de los coeficientes respectivos, las dos variables que recogen el comportamiento anterior del cliente respecto a las operaciones concedidas por la entidad presentan signo positivo, lo que significa que provocan un aumento en la puntuación obtenida a partir de la función discriminante, acercando así al individuo al grupo de clientes morosos. Todo lo contrario ocurre con la variable que recoge

la garantía hipotecaria aportada por el cliente, que presenta signo negativo, como era de esperar, ya que las entidades financieras exigen la presentación de garantías precisamente como medio para reducir el riesgo de impago. Por su parte, el hecho de que el cliente haya fijado recientemente su residencia en la zona provoca un aumento en la puntuación discriminante, como muestra el signo positivo del coeficiente asociado a esta variable, mientras que el estado civil tiene un efecto negativo, también previsible, pues esta variable de tipo cualitativo se ha introducido en el modelo a partir de variables ficticias, tomando como referencia la categoría de soltero, y cualquier otra modalidad de la variable provoca una reducción en la puntuación discriminante, con respecto a aquélla.

- Para la valoración de la significación de la función discriminante obtenida, el autovalor correspondiente, que explica el 100% de la varianza, y el coeficiente de correlación canónica, con un valor de 0'857, determinan una alta capacidad de discriminación de dicha función. Por otro lado, el test Lambda de Wilks, junto con su correspondiente estadístico χ^2 , que contrastan la hipótesis nula de igualdad de medias en los dos grupos, permiten rechazar dicha hipótesis, señalando la significación de la función en cuanto a la discriminación entre grupos.
- La capacidad predictiva de las funciones discriminantes se evalúa a partir de la matriz de clasificación que se muestra en el anexo A2, en la que se recogen los valores observados para la variable dependiente y los estimados por el modelo, mediante la determinación de la puntuación de corte óptima para la asignación de los casos a cada grupo definido por la variable dependiente. La puntuación de corte depende de la puntuación discriminante obtenida por los distintos individuos, que se calcula a partir de las ponderaciones asignadas a las variables incluidas en la función discriminante y de los valores que dichas variables presentan para cada caso.

Para el modelo correspondiente a la muestra de clientes del estudio se ha conseguido un porcentaje total de clasificación correcta de un 97'1%, que corresponde a una tasa de aciertos del 88'9% de clientes morosos y del 100% de clientes no morosos, tal como resulta de la tabla de clasificación recogida en el anexo A2.

La tabla 2 recoge todos los resultados obtenidos para el modelo estimado, respecto al grado de discriminación entre grupos y a la significación global de la función discriminante.

Tabla 2. Evaluación de la función discriminante

Autovalores				Significación global			
Autovalor	% varianza	% acumul.	Correlación canónica	Lambda de Wilks	χ^2	gl	Sig.
2,757	100	100	,857	,266	86,698	5	,000

4.3.2. Resultados del análisis de regresión logística

El modelo de regresión logística permite determinar la probabilidad de que un cliente sea moroso, mediante los coeficientes estimados correspondientes a cada variable explicativa incluida en el modelo y que reflejan la influencia de cada una de ellas en el fenómeno analizado.

Los criterios que el programa SPSS presenta como opciones para la selección de las variables explicativas, cuando se aplica el análisis de regresión logística, son el estadístico de Wald y el test de razón de verosimilitud ó χ^2 del modelo. El criterio utilizado en el estudio empírico ha sido éste último, ya que si se considera el test de Wald, el procedimiento finaliza tras la eliminación de la primera variable que se ha introducido, dando lugar al modelo inicial, que sólo contiene la constante.

- En razón del cambio o mejora que tiene lugar en el estadístico de razón de verosimilitud cuando se introduce cada variable, se obtiene un modelo final con cuatro variables independientes, por mostrar significatividad en la explicación de la variable dependiente, que se muestran, con sus respectivos coeficientes, en la tabla 3:

Tabla 3. Modelo de regresión logística estimado

Variable	β	E.T	Wald	gl	Sig.
NUMIMPAG	25,484	205,829	,015	1	,901
NUEVARES	14,929	712,545	,000	1	,983
TRASPASO	13,557	131,010	,011	1	,918
VEHICULO	11,765	131,007	,008	1	,928
Constante	-13,557	131,002	,011	1	,918

Las cuatro variables presentan coeficientes con signo positivo, indicando que cuando la variable independiente aumenta en una unidad, el logaritmo de la *odds* de la probabilidad de ser un cliente moroso aumenta en el valor del coeficiente respectivo. Además, el efecto de las variables sobre la probabilidad de que un cliente sea moroso es importante, dado que el valor de los coeficientes es muy alto.

Por otra parte, de las cuatro variables del modelo, dos se refieren a una de las características de la operación, en concreto al destino de la misma, ya que han resultado significativas las variables que recogen si la inversión se destina al traspaso de un negocio o a la compra de un vehículo. Otra de las variables hace referencia a una característica personal del cliente, si es de nueva residencia en la localidad. Por último, la variable número de impagos anteriores, ya hemos comentado que refleja la situación del individuo respecto a las operaciones que le ha concedido la entidad.

- Estimado el modelo a partir de los valores observados, se procede a evaluar su significación global y bondad de ajuste, ya que la significación individual de las variables independientes se ha valorado a la hora de decidir su inclusión o no en el modelo, en función de su contribución a la mejora del estadístico de razón de verosimilitud. Además de este estadístico, que es significativo, se calculan dos coeficientes similares al coeficiente de determinación R^2 del análisis de regresión lineal, que son el R^2 de Cox y Snell, y el de Nagelkerke, cuyos valores, próximos a uno, muestran que el ajuste realizado es bueno. Otro estadístico utilizado para valorar la bondad de ajuste del modelo es la prueba de Hosmer–Lemeshow, que resulta de la agrupación de las observaciones, para cada uno de los dos grupos definidos por la variable dependiente, en una tabla de contingencia, obteniéndose un resultado significativo.
- Un último aspecto a valorar es la eficacia o capacidad predictiva del modelo, que se realiza a partir de una tabla de clasificación, similar a la que se obtiene en el análisis discriminante y que se muestra, a su vez, en el anexo A2. En la tabla se recogen los valores observados de la variable dependiente, distinguiendo entre clientes morosos y no morosos, y los estimados por el modelo, que permiten clasificar a los individuos o clientes de la muestra en cada uno de los dos grupos, mediante el establecimiento de un punto de corte de 0'5 para las probabilidades predichas. El modelo estimado consigue una tasa de aciertos del 97'1%, de los cuales están correctamente clasificados un 98'1% de clientes no morosos y un 94'4% de clientes morosos. Todos los resultados obtenidos respecto a la valoración de la significación y bondad de ajuste del modelo se muestran en la tabla 4.

Tabla 4. Evaluación del modelo de regresión logística

Significación global			R cuadrado		Prueba de Hosmer y Lemeshow		
χ^2	gl	Sig.	Cox y Snell	Nagelkerke	H-L	gl	Sig.
71,292	4	,000	,639	,939	,000	3	1,000

3.4.3. Comparación de resultados

Una vez presentados los resultados obtenidos con cada técnica para la muestra de clientes objeto del estudio, efectuamos a continuación un análisis comparativo de los mismos.

Por lo que respecta a los modelos estimados, la función discriminante obtenida incluye cinco variables explicativas, dos de las cuales, relativas a la residencia del individuo y al número de impagos anteriores, también se contienen en el modelo obtenido mediante la técnica de regresión logística, pero a éstas se añaden otras dos, que hacen referencia al destino de la financiación solicitada. Por su parte, las tres variables que completan la función discriminante se refieren a la duración de los retrasos en el pago producidos con anterioridad, al estado civil del cliente y si éste ha aportado garantía real, ya sea de tipo hipotecario u otra.

Tabla 5. Comparación de modelos

Análisis discriminante		Regresión logística	
Variables	Ponderación	Variables	Ponderación
NUMIMPAG	1,013	NUMIMPAG	25,484
NUEVARES	,624	NUEVARES	14,929
DURARETR	,567	TRASPASO	13,557
GARHIPOT	-,502	VEHICULO	11,765
ESTCIVIL	-,440		
Tasa de aciertos		Tasa de aciertos	
Cientes morosos	88,9%	Cientes morosos	94,4%
Cientes no morosos	100%	Cientes no morosos	98,1%
Total	97,1%	Total	97,1%

Con tales variables, ambos modelos estimados presentan la misma capacidad predictiva, consiguiendo una tasa de aciertos global del 97'1% de clientes de la muestra, si bien, dicho porcentaje se alcanza de distinta forma por cada modelo. Así, mientras el análisis discriminante permite clasificar correctamente a la totalidad de clientes no morosos, frente al 98'1% que consigue el modelo de regresión logística, éste consigue un porcentaje de clasificación correcta de clientes morosos mayor que la función discriminante (94'4% frente a 88'9%).

Por tanto, podemos observar que tanto el análisis discriminante como la regresión logística consiguen un mayor porcentaje de aciertos entre los clientes no morosos que entre los individuos morosos, llegando incluso el primero a clasificar correctamente la totalidad de

dichos clientes, porcentajes que descienden respecto a los sujetos morosos, sobre todo en el análisis discriminante, aunque ambas técnicas poseen la misma eficacia predictiva total.

4. CONCLUSIONES

- Los dos modelos estimados mediante análisis discriminante y regresión logística son significativos, como lo muestran los estadísticos calculados para valorar la significación de cada modelo. Además, poseen una elevada capacidad predictiva, sobre todo entre los clientes no morosos, ya que la primera técnica clasifica correctamente a la totalidad de tales individuos, alcanzando, la segunda, una tasa de aciertos del 98'1%. Dichos porcentajes son menores en cuanto a los individuos morosos, siendo los porcentajes de clasificación correcta del 88'9% para el análisis discriminante y del 94'4% para la regresión *logit*, lo que resulta en una tasa de aciertos del 97'1% para ambos modelos.

- En cuanto a las variables que determinan que un cliente sea moroso, han resultado significativas tanto aquéllas relativas a las características de la operación formalizada como las que reflejan el perfil del individuo. Así, dos variables, referidas al número de impagos anteriores y a la residencia del cliente y pertenecientes, por tanto, a cada uno de los dos bloques de variables explicativas seleccionadas, se incluyen en los dos modelos estimados. A éstas se añaden, para completar el modelo *logit*, otras dos variables, que recogen si la inversión financiada se ha destinado al traspaso de un negocio o a la adquisición de un vehículo. Por su parte, la función discriminante está formada por tres variables más, que hacen referencia a la duración de los retrasos que el cliente haya tenido, a la aportación de garantía hipotecaria y al estado civil del individuo.

- Todas las variables incluidas en el modelo de regresión logística tienen un efecto positivo en la variable dependiente, dado el signo de los coeficientes asociados, por lo que un cambio de una unidad en la variable provoca un aumento en el logaritmo de la *odds* de la probabilidad de que un cliente sea moroso, en el valor del coeficiente respectivo. Por lo que respecta a la función discriminante, las variables relativas al historial de impagos del cliente y la que recoge si es de nueva residencia en la zona provocan aumentos en la puntuación obtenida a partir de la función discriminante, mientras que la aportación de garantía real y el estado civil del cliente, respecto a la situación de soltería, disminuyen el valor de la puntuación discriminante.

La conclusión global que se extrae es que tanto la técnica de análisis discriminante como la de regresión logística son adecuadas para el estudio y predicción de la morosidad, consiguiéndose mediante ambas técnicas una elevada eficacia predictiva. Esto es así porque las condiciones exigidas para la aplicación del análisis discriminante, en cuanto a la igualdad de varianzas en los dos grupos definidos por la variable dependiente y la normalidad de las variables independientes en la población de la que procede la muestra, se verifican en la muestra de clientes objeto del estudio. Por ello, los resultados obtenidos mediante esa técnica son similares a los conseguidos mediante regresión logística, que es el método alternativo aplicado para el análisis de variables cualitativas cuando se incumplen tales restricciones.

5. BIBLIOGRAFÍA

- ABAD, F.; VARGAS, M. (2002) *Análisis de datos para las Ciencias Sociales con SPSS*. Granada: Ed. José Carlos Urbano Delgado, S.L.
- ALONSO MARTÍNEZ, C. (2001) "Credit scoring y consentimiento en la protección de datos", *Perspectivas del sistema financiero*, nº 72, pp. 85-91.
- ALTMAN, E.I.; SAUNDERS, A. (1998) "Credit risk measurement: Developments over the last 20 years", *Journal of Banking & Finance*, vol. 21, pp. 1721-1742.
- BANEGAS GARCÍA, C.; GARCÍA MARTÍNEZ, F. (2000) "El riesgo de crédito en la banca ante la nueva regulación del Banco de España", *Banca & Finanzas*, nº 60, pp. 6-9.
- BERTRAN JORDANA, J. (1994) "Control y gestión de los riesgos financieros", *Alta Dirección*, nº 173, pp. 65-72.
- BOAL VELASCO, N.; GONZÁLEZ SÁNCHEZ, M. (2001) "Estimación del riesgo de crédito mediante modelos internos", *Banca & Finanzas*, nº 66, pp. 40-45.
- BOYES, W.J.; HOFFMAN, D.L.; LOW, S.A. (1989) "An Econometric Analysis of the Bank Credit Scoring Problem", *Journal of Econometrics*, nº 40, pp. 3-14.
- CEA D'ANCONA, M.A. (2002) *Análisis multivariable. Teoría y práctica en la investigación social*. Madrid: Ed. Síntesis.
- CRUZ GONZÁLEZ, F.J. de la (1998) "Enfoques cuantitativos para la predicción del riesgo de crédito", en CALVO-FLORES SEGURA, A. y GARCÍA PÉREZ DE LEMA, D. (coord.) *Predicción de la insolvencia empresarial*. Madrid: Monografías AECA.

- DOLDÁN TIÉ, F.; RODRÍGUEZ LÓPEZ, M. (2002) *La Gestión del Riesgo de Crédito. Métodos y modelos de Predicción de la Insolvencia Empresarial*. Madrid: Monografías AECA.
- GABÁS TRIGO, F. (1997) "Predicción de la insolvencia empresarial", en CALVO-FLORES SEGURA, A. y GARCÍA PÉREZ DE LEMA, D. (coord.): *Predicción de la insolvencia empresarial*. Madrid: Monografías AECA.
- GARCÍA, D.; ARQUES, A.; CALVO-FLORES, A. (1995) "Un modelo discriminante para evaluar el riesgo bancario", *Revista Española de Financiación y Contabilidad*, nº 82, pp. 175-200.
- GONZÁLEZ SÁNCHEZ, M. (2000) "Nuevas tendencias en la gestión de riesgos: riesgo de crédito", *Perspectivas del sistema financiero*, nº 69, pp. 59-92.
- HAND, D.J. (2001) "Modelling consumer credit risk", *IMA Journal of Management Mathematics*, vol.12, nº 2, pp.139-155.
- ; HENLEY, W.E. (1997) "Statistical Classification Methods in Consumer Credit Scoring: a Review", *Journal of the Royal Statistical Society, Series A*, vol. 160, nº 3, pp. 523-541.
- LAFFARGA BRIONES, J.; MARTÍN MARÍN, J.L.; VÁZQUEZ CUETO, M.J. (1985) "El análisis de la solvencia en las instituciones bancarias: Propuesta de una metodología y aplicaciones a la Banca española", *ESIC-MARKET*, nº 48, pp. 51-73.
- LAITINEN, E.K. (1999) "Predicting a corporate credit analyst's risk estimate by logistic and linear models", *International Review of Financial Analysis*, vol. 8, nº 2, pp. 97-121.
- LOPEZ, J.A.; SAIDENBERG, M.R. (2000) "Evaluating credit risk models", *Journal of Banking & Finance*, vol. 24, nº 1-2, pp. 151-165.
- LUQUE MARTÍNEZ, T. (coord.) (2000) *Técnicas de análisis de datos en investigación de mercados*. Madrid: Ediciones Pirámide.
- PALEPU, K.G. (1986) "Predicting takeover targets. A methodological and empirical analysis", *Journal of Accounting and Economics*, vol. 8, pp. 3-35.
- PEÑA, D. (2002) *Análisis de datos multivariantes*. Madrid: Editorial McGraw-Hill.
- RAMÍREZ COMEIG, I. (1998) "Determinantes de la insolvencia en las operaciones avaladas por las sociedades de garantía recíproca: una aplicación del análisis discriminante y del análisis logit", *Revista Europea de Dirección y Economía de la Empresa*, vol. 7, nº 1, pp. 149-166.
- RODRÍGUEZ FERNÁNDEZ, J.M. (1987) "Crisis en los bancos privados españoles: un modelo logit", *Investigaciones Económicas*, suplemento, pp. 59-64.

- (1989) "Análisis de las insolvencias bancarias en España: un modelo empírico", *Moneda y Crédito*, nº 189, pp. 187-227.
- RUIZ, G.; JIMÉNEZ, J.I.; TORRES, J.J. (2000) *La gestión del riesgo financiero*. Madrid: Ediciones Pirámide.
- SALAS, V.; SAURINA, J. (2002) "Credit Risk in Two Institutional Regimes: Spanish Commercial and Savings Banks", *Journal of Financial Services Research*, vol. 22, nº 3, pp. 203-224.
- SANTOS PEÑA, J.; MUÑOZ ALAMILLOS, A.; JUEZ MARTEL, P.; GUZMÁN JUSTICIA, L. (1999) *Diseño y tratamiento estadístico de encuestas para estudios de mercado*. Madrid: Editorial Centro de Estudios Ramón Areces, S.A.
- SAURINA SALAS, J. (1999) "Determinantes de la morosidad de las Cajas de Ahorros españolas", *Investigaciones Económicas*, vol. XXII, nº 3, pp. 393-426.
- SOLER, M.; MIRÒ, A. (2001) "Enfoques cuantitativos para el riesgo de crédito de particulares y su aplicación a realidades nacionales diferentes", *Perspectivas del sistema financiero*, nº 72, pp. 43-51.
- THOMAS, L.C. (2000) "A Survey of Credit and Behavioural Scoring: Forecasting financial risk of lending to consumers", *International Journal of Forecasting*, vol. 16, nº 2, pp. 149-172.
- TOMÁS, J., AMAT, O. y ESTEVE, M. (2002) *Cómo analizan las entidades financieras a sus clientes*. Barcelona: Gestión 2000.com, 2º edición.
- WIGINTON, J.C. (1980) "A Note on the Comparison of Logit and Discriminant Models of Consumer Credit Behaviour", *Journal of Financial and Quantitative Analysis*, vol. XV, nº 3, pp. 757-770.
- WILSON, N.; SUMMERS, S.; HOPE, R. (2000) "Using Payment Behaviour Data for Credit Risk Modelling", *International Journal of the Economics of Business*, vol. 7, nº 3, pp. 333-346.

6. ANEXO

Anexo A1. VARIABLES

Variable	Categorías
individu	Número de individuo
morosida	1, si el cliente es moroso; 0, si no lo es
importe	Importe del crédito, en miles de euros
plazo	Duración de la operación, en meses
vivienda	1, si el destino de la inversión es la compra o reforma de vivienda; 0, en otro caso
vehiculo	1, si el destino es la compra de vehículos; 0, en otro caso
activida	1, si el destino es la compra de locales comerciales o activos para actividad profesional; 0, en otro caso
canceledu	1, si el destino es la cancelación de deudas; 0, en otro caso
traspaso	1, si el destino de la inversión es mercantil; 0, en otro caso
otrgasto	1, si el destino son otros gastos o compras; 0, en otro caso
proceden	1, si es por iniciativa del cliente; 2, de la entidad; 3, campaña de marketing; 4, otra persona física o jurídica; 5, otra entidad
tipointe	1, si fijo; 0, si es variable
garperso	1, si se aporta garantía personal; 0, en otro caso
garhipot	1, si se aporta garantía hipotecaria o real; 0, en otro caso
garaval	1, si se aporta aval; 0, en otro caso
edad	Edad del cliente, en años
estcivil	1, si es soltero; 2, casado/pareja; 3, separado/divorciado; 4, viudo
cargafam	Número de personas a cargo del cliente
slaboral	1, por cuenta propia; 2, por cuenta ajena funcionario; 3, por cuenta ajena indefinido; 4, por cuenta ajena eventual; 5, por cuenta propia y ajena funcionario; 6, por cuenta propia y ajena indefinido; 7, por cuenta propia y ajena eventual; 8, jubilado/prejubilado; 9, desempleado
ingresos	Ingresos medios mensuales, en euros
antigued	Antigüedad en el trabajo, en años
propviv	1, si posee vivienda sin cargas; 2, posee vivienda con cargas; 3 no posee vivienda
nuevares	1, si es de nueva residencia; 0, en otro caso
nuevodi	1, si es nuevo cliente; 0, en otro caso
compensa	1, si tiene compensaciones; 0, en otro caso
numectas	Número de cuentas del cliente
otrasope	1, si tiene otras operaciones activas; 0, en otro caso
numerope	Número total de operaciones
imptotop	Importe total de las operaciones activas, en miles de euros
numimpag	Número de impagos anteriores
duraretr	Duración máxima del retraso, en meses

Anexo A2. Capacidad predictiva

		Pronosticado		Porcentaje global	
		MOROSIDAD			
Observado		no	sí		
Análisis discriminante	MOROSIDAD	no	52	0	100,0000
		sí	2	16	88,8889
				Porcentaje global	97,1429
Regresión logística	MOROSIDAD	no	51	1	98,0769
		sí	1	17	94,4444
				Porcentaje global	97,1429